ALGORTIHMS

Detection of similarity between documents with hashing

Roger Vilaseca DarnÃ©, Xavier Lacasa Curto and Xavier MartÃn Ballesteros

10th December 2018

Contents

1 Introduction

References

[1] Anand Rajaraman, Jure Leskovec and Jeffrey D. Ullman. *Mining of Massive Datasets*. Cambridge University Press. (December 30, 2011).

[2] cmhteixeira. *Locality Sensitive Hashing (LSH)* [online]. (November 29, 2017). <https://aerodatablog.wordpress.com/2017/11/29/locality-sensitive-hashing-lsh/>[Consulted: December 12, 2018].

[3] Hubert Brylkowski. *Locality sensitive hashing − LSH explained* [online]. (October 6, 2017). <https://medium.com/engineering-brainly/locality-sensitive-hashing-explained-304eb39291e4>[Consulted: December 15, 2018].

[4] Jeffrey D. Ullman[Mining Massive Datasets]. (July 23, 2016). *3 2 Minhashing 25 18*. <https://www.youtube.com/watch?v=96WOGPUgMfw>.

[5] Jeffrey D. Ullman[Mining Massive Datasets]. (July 23, 2016). *3 3 Locality Sensitive Hashing 19 24*. <https://www.youtube.com/watch?v=_1D35bN95Go>.

[6] santhoshhari. *Locality Sensitive Hashing: Application of Locality Sensitive Hashing to Audio Fingerprinting* [online]. (n. d.). <https://santhoshhari.github.io/Locality-Sensitive-Hashing/>[Consulted: December 16, 2018].

[7] Shikhar Gupta. *Locality Sensitive Hashing: An effective way of reducing the dimensionality of your data* [online]. (June 29, 2018). <https://towardsdatascience.com/understanding-locality-sensitive-hashing-49f6d1f6134>[Consulted: December 16, 2018]